Gaming Systems Test Plan

TASK 3: Develop Energy Performance Measurements & Benchmark Testing Options

December 2, 2016

A Plug-loads Game Changer: Gaming System Energy Efficiency without Performance Compromises - CEC/LBNL Agreement # EPC-15-023

Task 3 - Gaming Systems Test Plan

CEC/LBNL Agreement # EPC-15-023

Introduction

In accordance with the *EPC-15-023 Exhibit A - Attachment 1 - Schedule*, this Test Plan provides a concise summary of the structure, rationale and schedule for the physical hardware and software testing that will occur under the *Task 5: Identify Options to Improve Energy Performance of Gaming PC Systems* work scope.

This document has been developed and written as both the above listed EPIC project deliverable and as a LBNL internal 'living' document, during the exploration and discovery phases of Tasks 3, 4 and 5. This document provides a basic overview of the schematic logic and testing methods that we will be evaluated and refined during the Task 5 Testing Lab setup and commissioning period from November 2016 until the beginning of formal product testing in 2017. The performance metrics, measurement methods and testing procedures outlined in this report will be codified into final testing scripts prior to formal testing and the resulting standardized procedures will be documented in the Draft and Final versions of the Task 3 and 4 shared project Deliverable, the *Gaming Systems Energy Performance Measurements & Benchmark Testing Procedures Report*.

Testing Scope and Goals

The general testing plan of this project is to measure the Unit Energy Consumption (UEC) and User Experience Performance (UXP) of a set of gaming systems hardware using appropriate benchmarking software. The gaming systems hardware will include a selection of market representative Gaming Personal Computers (GPC), Video Game Consoles (VGC), and TV Game Devices (TVGD) identified by Task 2 Market Research. The benchmarking software will include game titles identified by Task 2 and synthetic benchmarks identified by Task 3.

It's important to note that the UEC and UXP metrics are equal and essential parts of this test plan since the project goal is to identify energy efficiency opportunities without compromising the system performance. UEC measurement is straightforward and is similar to that done for numerous other consumer products. But the whole point of gaming devices is to deliver fun and entertainment, qualities that are extremely hard to quantify. Any successful energy savings strategy will need to prevent even the perception of a performance sacrifice if it is to succeed. With this goal in mind a number of parameters will be measured in order to determine a reasonable UXP metric that can track any gaming user detected performance degradations.

The general progression of the Task 3 developed testing regime is as follows.

1. Measure the system baseline UEC and UXP for:
    - Entry-level GPCs

    - Mid-level GPCs
    - High-end GPCs
    - Laptop GPCs
    - Video Game Consoles
    - TV Gaming Devices
2. Develop a set of energy efficiency measures (hardware, firmware, and software) to be implemented on each system
3. Measure any resulting UEC and UXP changes in order to inform potential technical savings and/or and potential cross platform benchmarking methodologies for energy end-use tracking at the state scale

Planned Schedule

The planned schedule for Task 5 hardware and software testing includes six distinct subtasks spanning from September 2016 to the completion of testing in May 2017.

- Produce Gaming System Acquisition Specifications Report (10/03/16 - 12/06/16)
- Test Bed and Lab Development and Setup (09/05/16 - 01/24/17)
- Specify and Procure all Test Gaming Systems and peripherals (10/17/16 - 01/24/17)
- Implement Test Plan (begin testing 01/24/17)
- Develop and implement energy efficiency measure modifications (03/13/17 - 05/05/17)
- Implement key non-energy test procedures (01/25/17 - 05/22/17)

More details on the planned schedule for the above listed milestones are provided in the following image capture of the *Task 5: Identify Options to Improve Energy Performance of Gaming PC Systems* section of our internal project management tool (developed in SmartSheet). During the writing of this Test Plan in the project schedule timeline, we are firmly in the process of tooling up the new testing lab and beginning the test equipment commissioning process.

Task Name	2016				2017			
	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
Identify Energy Perf. Options (Testing)								
Produce Gaming System Acquisition Specifications Report to include, but not limited to a summary of GPC system guidelines and equipment energy intensities								
Test Bed and Lab Development and Setup								
Develop Test Bed Specifications								
Room Setup - Test Lab Area								
Installation and Setup of Test Equipment								
Team authorize final test lab setup								
Specify and Procure all Test Gaming Systems and associated peripheral equipment								
Begin Specify test systems (GPC and Video Gaming) to be procured								
Order all test systems that will not change in final vendor report								
Order balance of specified systems as defined in vendor report								
All test systems and peripherals procured								
Implement Test Plan								
Begin Testing and logging of results								
Conduct baseline energy performance benchmark tests on all GPCs test systems								
Conduct Cross-platform testing								
Conduct sensitivity tests								
Conduct nameplate vs in-use power requirements tests								
Develop and implement energy efficiency measure modifications for all GPC test systems and repeat energy performance benchmark testing in a stepwise fashion								
Perform an economic (cost-benefit) assessment of the identified efficiency improvements								
Implement key non-energy test procedures								
Determine available budget for testing scope of non-energy metrics (e.g. internal temperatures, noise levels and thermal comfort)								
Begin non-energy metrics testing and logging of results								
Conduct non-energy metrics testing								
Testing Complete								

Planned System Tests & Measurements

Already mentioned above are the two primary system measurements required for this project in order to calculate a system efficiency metric: Unit Energy Consumption (UEC) and User Experience Performance (UXP). UEC will be measured as the system energy consumed, in watt-hours units. We haven't yet determined an appropriate unit combination that encompasses all the UXP factors under consideration.

In addition to the primary measurements (UEC and UXP), we will also measure a number of related parameters that can provide non-energy benefits to the gaming experience. These include component temperatures, which can often correspond to heat loads being imposed into

the user's home by the gaming system leading to potential human thermal discomfort, and excessive system operation noise which can contribute to acoustical discomfort.

Energy Consumption

Energy consumption will be measured at both the system and component levels. System power is supplied by the mains (120VAC) and includes power used by both the Unit Under Test (UUT, i.e. GPC, VGC or TVGD) and the display that the UUT is driving (both conventional and Virtual Reality). Component power consists of low-voltage (3.3, 5, or 12 VDC) and current measurements of both internal components (e.g. CPU, GPU, etc) and external components (e.g. speakers, tracking cameras, etc.).

System Power Measurement

System power will be measured using a Chroma 66202 digital power meter[1] using a A662003 measurement test fixture. One-second interval readings of power (in watts) and power factor will be recorded using a USB connection to a test bench PC running the Chroma Softpanel software application.

Component Power Measurement

Component power will be measured using a Measurement Computing USB-1608FS-Plus data acquisition system[2] connected to a test bench PC running the DASYLab software application. Internal components can be powered by up to three different DC voltage rails (3.3, 5, and 12 volt). Each voltage will be measured either directly or using a voltage divider (for 12 volt rails). Each current will be measured using a Pico TA234 DC current clamp. One-second values of voltage, current (in amps), and power (V x A, in watts) will be recorded using the DASYLab software.

User Experience Factors

The most popular user experience parameter used by the gaming industry is frame rate (the rate at which image frames are rendered to the display screen). After a long period of generally improved gaming hardware generation cycles delivering ever increasing frame rates, the emphasis is now potentially declining as systems are typically able to deliver frame rates above those perceivable by most gamers. Other aspects of frame rate that affect user experience include jitter, tearing, dropped frames, and frame time variance or smoothness.

Image Quality Metrics & Measurement Tools

Image quality determination typically involves many factors which occur at multiple measurement points within the system image "rendering pipeline." The rendering process

[1] http://www.chromausa.com/product/single-channel-digital-power-meters-66201-66202/
[2] http://www.mccdaq.com/usb-data-acquisition/USB-1608FS-Plus.aspx

essentially begins at the gaming system Processor Units - often simply referred to as the "Gaming Engine" - and out to the actual display hardware. The Gaming Engine is a software process which cuts across various hardware and accessory component lines, drawing on the CPU and GPU processing resources to render each image frame in preparation for steaming the video data to the display. The following list of metrics will be used to track the image quality performance:

- Frame Rate
  - The average measured frame rate, reported in Frames per Second (FPS)
  - Frame Time - The length of time (in milliseconds) it takes the graphics card to render and display each individual frame
  - Frame Time Statistics - Minimum, Maximum, Mean, Quartiles, Standard Deviation.
  - Frame Time Variance - Represented in milliseconds, this indicates the frame-to-frame smoothness quality (also referred to as Micro Stuttering[3])
- Vertical Sync
  - Input Lag[4] - also called Scanout, a condition where the display and system refresh rates are out of sync and frames are cut off before being completed (also know as Frame Runts)
  - V-Sync Stutter (causes inconsistent on screen motion and image tearing)

Frame rates and quality will be measured at critical points in the image processing path using the following tools:

- FRAPS - A software application that measures the frame rate of the gaming engine rendering output[5]
  - Produces average frame rate = total frames / total time.
  - Compatible with DX11 or lower
  - Only runs on Windows GPC systems
- PresentMon - A software application that measures the frame rate of the gaming engine rendering output[6]
  - Measures Universal Windows Platform (UWP) and DX10, 11, and 12 games
  - Only runs on Windows GPC systems
- FCAT (Frame Capture Analysis Tool) - A combination of video capture hardware and software tools that samples and analyzes all frames leaving the graphics card output to the display input[7]
  - Produces frame rates and statistics over time (once per second)
  - Provides logging of frame quality factors (runts, tearing, etc.)
  - Available for all GPC systems, plus potentially VGCs and TVGDs

[3] https://en.wikipedia.org/wiki/Micro_stuttering
[4] http://www.anandtech.com/show/2803/5
[5] http://www.fraps.com
[6] https://github.com/GameTechDev/PresentMon
[7] http://www.geforce.com/hardware/technology/fcat

Temperature

Temperature is a limiting design factor for critical electronic components such as the CPU and GPU and is an indirect indicator of heat load stresses on the system. While there are many direct methods of measuring component temperatures, such as attaching thermocouples or using infrared video devices, these are time consuming to set up and too costly for this project. In general these spot temperature readings have a complicated indirect relationship to overall system heat load, since it is highly affected by the specific heat transfer coefficient and contact area for each component being evaluated. Given the complication of the UA calculations and the secondary nature of this measurement to the project goals, we will instead use readily available system software tools to read data from internal component sensors.

CPU temperatures will be measured using CoreTemp[8] software, which reads directly from the Digital Thermal Sensor (DTS) located in each individual processing core. GPU temperature will be measured using the GPU-Z[9] software package.

Sound

System noise can be generated by hard drive units, optical disk drives and “coil whine” from some electronic components, however the primary source of noise in gaming systems comes from the various cooling systems necessary to maintain acceptable component temperatures. As gaming systems have gotten more powerful with higher component density, heat loads have increased and active cooling is now required in almost all cases. More efficient systems typically require less cooling and therefore can be quieter, thereby providing potentially valuable non-energy benefit to the user.

Since this project does not have the budget for a comprehensive acoustical measurement lab and due to the secondary nature of this performance metric, we will conduct simple qualitative measurements of the ambient sound levels occurring in our regular lab during Active Mode tests. The system sound level will be measured using an Extech 407750[10] sound level meter. The meter will be positioned at the height of the UUT and at a standardized distance for all tests (to be determined after the Test Lab setup is complete). Measurements will be done using dB(A) weighting and recorded on a one-second interval basis using a PC interface.

Planned Testing Procedures

The determination of a viable active mode gaming system energy and performance test procedure is a key outcome of Task 5. A robust and well designed test procedure should have the following features:

- Repeatable (Returns a consistent result)
- Representative (Of actual energy use)

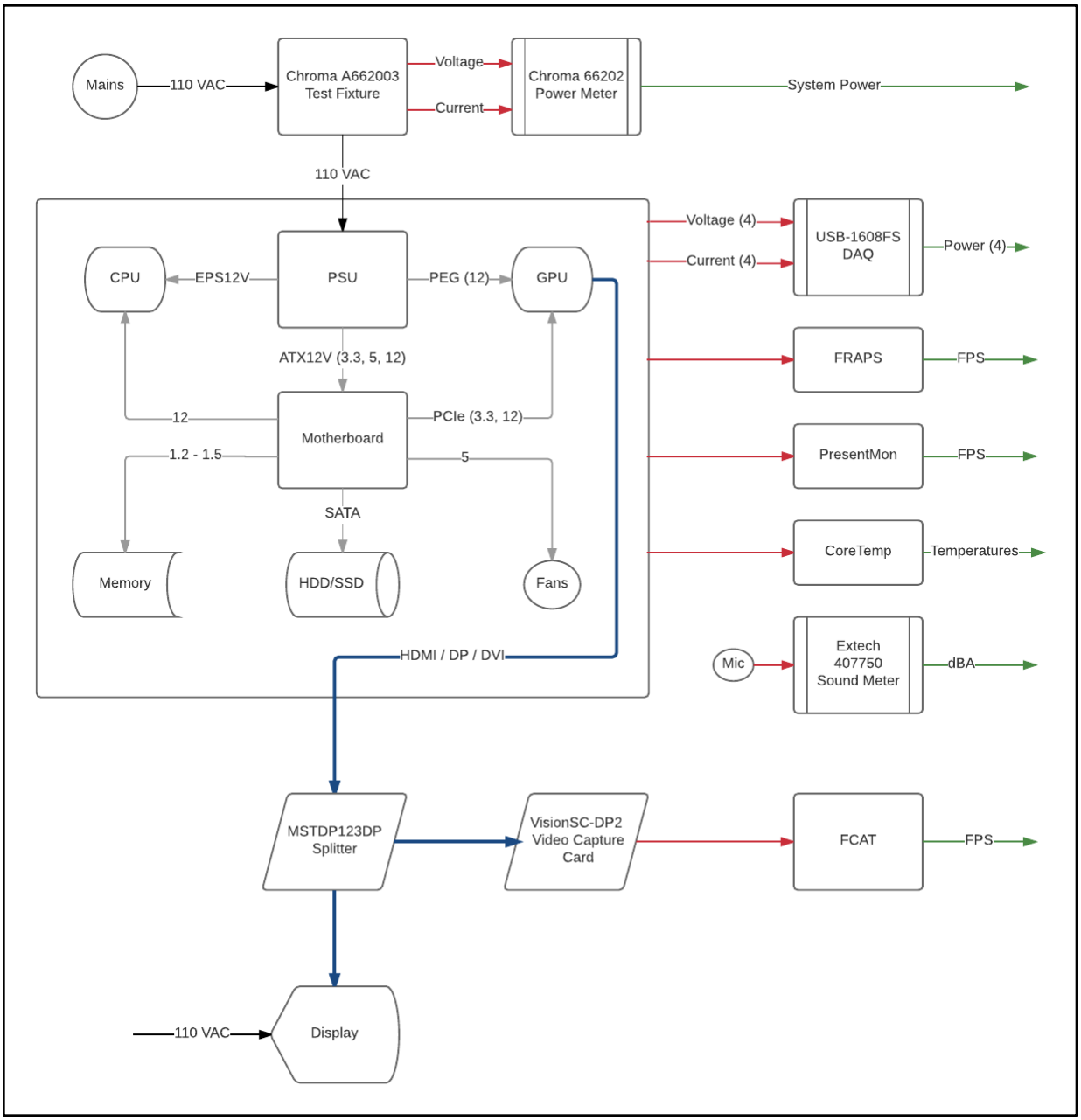
[8] http://www.alcpu.com/CoreTemp/
[9] https://www.techpowerup.com/gpuz/
[10] Extech 407750: Sound Level Meter with PC Interfac, http://www.extech.com/display/?id=14224

- Comparable (Across other product types)
- Open (Public disclosure)
- Stable (Over time)
- Neutral (Across market products)

All the GPCs, VGCs, and TVGDs ("Systems") will be tested using a fixed set of game titles and Synthetic Benchmarks, under project standardized test procedures in order to determine system power use and performance. The planned tests are listed in the following table, along with an estimate of the total number of tests that will be conducted.

Test	**Systems**	**Benchmarks**	**Total Tests**
Initial Exploratory Tests	4	14	56
Baseline GPC	16	8	128
Baseline VGC	10	6	60
Hardware Sensitivity	28	5	140
Software Sensitivity	4	30	120
Display Configuration	17	7	119
Non-active Mode	26	6	156
Efficiency Measures	36	5	180
Nameplate Power	16	4	64
TOTAL			**1,023**

The following diagram provides an alternative high level view of the current testing plan. The diagram documents the various power and data flows at a schematic level. The series of green arrows on the right side identify the data that will be logged and available for subsequent analysis in the project.



Initial Exploratory Tests

There are two main categories of GPC Benchmarks available: Games which have a Benchmark Mode feature and Synthetic Benchmarks[11]. PC games often have a Benchmark Mode, a feature that runs the game engine through a set of repeatable command tasks for a fixed period of time. The feature is often used as a convenient and robust method to test the suitability of a user's PC system for running the game. An additional, and possibly more common, use of this feature is the large gaming system product reviewer community for marketing purposes. Synthetic

[11] https://en.wikipedia.org/wiki/Benchmark_(computing)

benchmarks are written using many of the same techniques and software tools as actual games, but are designed to stress a system in an unbiased and repeatable way.

VGC and TVGD games do not appear to have benchmarking modes, likely because there is not the same user interest in system-to-system comparisons as there are for GPCs. In addition, VGCs are all closed hardware systems, so there are no Synthetic Benchmark products available for these platforms. The two options available for an active mode "benchmark" of these platforms are: 1) The tester does their best to run a game through a prescribed set of moves in order to minimize variations between sample runs, and 2) Placing a game in "pause" mode[12].

This series of Initial Exploratory Tests will evaluate market available Benchmark products with respect to their suitability as viable proxies of a valid gaming load on the systems being tested. The core goal will be to identify a single or shortlist of Synthetic Benchmarks that best replicate the typical system load profiles that we observe in our set of 10 selected game titles (identified in Task 4 work). The following long list of Synthetic Benchmarks will be screened for suitability of exploration testing, with approximately 4 being chosen for actual testing.

Name	Publisher	URL	Cost
Heaven or *Valley*	Unigine	https://unigine.com/products/benchmarks/heaven/	Basic - $0, Advanced - $15, Pro - $495
3DMark	FutureMark	http://www.futuremark.com/benchmarks/3dmark	Single-Seat License from $1495
Cinebench	Maxon	https://www.maxon.net/en/products/cinebench/	Free
Furmark	Ozone3D	http://www.ozone3d.net/benchmarks/fur/	Freeware
Novabench	Novawave Inc	https://novabench.com/	Free
Catzilla	AllBenchmark	http://www.catzilla.com/	Free,,Basic/Advanced - $99, Pro - $895
GFXBench	Kishonti Ltd.	https://gfxbench.com/benchmark.jsp	Call for price quote
Aida64	Final Wire	https://www.aida64.com/products/aida64-engineer	Engineer - $199
Geek Bench 4	Primate Labs	http://geekbench.com/	Pro - $80
Performance Test	Passmark	http://www.passmark.com/products/pt.htm	$27

[12] As suggested by Kieren Mayers, Director of Environment and Technical Compliance, Sony Computer Entertainment Europe Ltd., during a Video Game Console industry meeting at LBNL with the project team 11/18/16.

Baseline System Testing

Using the set of synthetic benchmarks and games titles determined by the Initial Exploratory Tests, this phase of testing will baseline the active mode UEC and UXP for all gaming systems being tested. 3 "representative" games, 3 active mode games (different set for each of the four system categories) and 2 synthetic benchmarks will be run on each of the test systems.

Hardware Sensitivity Testing

The hardware sensitivity tests look at the effect of different hardware choices and configurations on active mode energy use and performance. Because VGCs and TVGDs are closed systems and do not provide access to internal components, these tests will only be performed on GPCs. For each system configuration, a set of 5 benchmarks (3 games and 2 synthetic benchmarks) will be run. Hardware components tested will include CPU, GPU, RAM, storage, and power supply. A total of 28 configurations will be tested as follows:

Component	**Options**
CPU	Overclocking(3), Cooling(4)
GPU	Overclocking(3)
RAM	Type(3), Size(3), Speed(3)
Storage	HDD, SDD
Power Supply	Size(3), Efficiency(5)

Software Sensitivity Testing

The software sensitivity tests look at the effect of different software settings on the active mode energy and performance of four GPCs (one each from the entry-level, midrange, high-end, and laptop categories). The 30 software settings including game quality settings, OS settings, and BIOS settings are summarized below:

Setting	Options
Quality	Low, medium, high, ultra
Anti Aliasing	None, SSAA, MSAA, CSAA, TXAA, FXAA
Ambient Occlusion	None, SSAO, SSDO, HBAO
Filtering	None, Bilinear, Trilinear, Anisotropic
Other	Blur, DOF, Bloom, Adaptive Vsync, HDR
API	DX11, DX12, Vulkan
OS Settings	Not yet established
BIOS Settings	Not yet established

Display Configuration Testing

The display configuration tests look at the effect that display type, resolution and options have on active mode energy use and performance. Not only do displays use energy directly, they can have a significant effect on system energy use depending on their resolution and frame-rate requirements. For example, virtual reality (VR) displays may use less energy themselves due to their small size, but they can require significantly higher frame-rates, which may result in higher total energy use due to extra GPU energy consumption. Because they have been included in the EnergyStar program for many years, conventional computer monitor energy use is very well characterized. However, the energy use implications of new technologies such as VR, High Dynamic Range (HDR), and vertical sync control are still unknown. The 22 display configurations summarized in the table below will be tested using 5 benchmarks.

Configuration	Options
Resolution	720, 1080, 1440, 2160
Video connection	HDMI 1.4, HDMI 2.0, DisplayPort 1.2, DisplayPort 1.3, DVI-D
Number of displays	1, 2, 3
Virtual Reality	HTC Vive, Oculus Rift, Playstation VR
Vertical sync control	Off, On
Overclocking	60Hz, 75Hz, 144Hz

Non-Active Mode Testing

While we expect the majority of gaming system energy use to be in active mode, non-active mode energy use is still significant and must be accounted for in a representative way. Many standard non-active mode tests have been developed for both GPCs and VGCs, including:

- IEC 62623 - Desktop and notebook computers - Measurement of energy consumption[13]
- ENERGY STAR® Program Requirements for Computers v6.1[14]
- ENERGY STAR Game Console Version 1.0 Recognition Program[15]
- European Union Self-Regulatory Initiative[16]

For GPCs we will use the ENERGY STAR v6.1 test method, which has the following modes: Video streaming, Web browsing, Short idle, Long idle, Sleep, and Off. For the VGCs and TVGDs we will use the EU Self-Regulatory Initiative which has the following modes: Navigation, Media playback, Streaming, and Standby. In addition we will use a "Real-world" idle mode test based on the method developed by XERGY for determining a real-world adjustment factor[17]. Energy use for all 26 system configurations will be tested in the above modes.

Efficiency Measures Testing

The Task 5 work scope and the Task 9 Hack-A-Thon proceedings will be used to develop a set of hardware and software efficiency measures which will be fully applied to the 9 modifiable GPC systems (not all measures will be applicable to the laptops or Apple systems). The efficiency measures testing will look at the effect of the modified software and hardware combination measures on the active mode UEC and UXP for each system. Each of the 9 systems will be tested in four configurations using up to five benchmarks.

Nameplate Power Testing

Some of the major components of gaming systems have a rated or "Nameplate" power draw. For CPUs and GPUs this is the thermal design power (TDP) specification - defined as the maximum amount of heat that must be dissipated under normal operation. This rating is most often used to determine the size of the power supply unit (PSU) for a GPC and misunderstandings can often lead to under or oversizing of PSU when designing a do-it-yourself GPC. 9 CPUs and 7 GPUs will be tested in two modes (idle and active).

---

[13] https://webstore.iec.ch/publication/7271
[14] https://www.energystar.gov/sites/default/files/Version%206.1%20Computers%20Program%20Requirements%20%28Rev.March-2016%29_0.pdf
[15] https://www.energystar.gov/products/spec/game_console_version_1_0_recognition_program_pd
[16] http://efficientgaming.eu/fileadmin/user_upload/Games_Consoles_Self-Regulatory_Initiative.pdf
[17] http://docketpublic.energy.ca.gov/PublicDocuments/14-AAER-02/TN211731_20160606T163325_California_Investor_Owned_Utilities_Comments_California_Investo.pdf

Testing Organization & Results Logging

CA Gaming Systems - Test Procedure Input Sheet

Description	
Notes	
Technician	Jimmy Mai
Gamer ID	
Date	11/30/2016
LETS Hours	1.2

	Description	ID
System	Alienware Alpha Base	E1
Display	Oculus Rift	VR1
Benchmark	3DMark Fire Strike	B2
OS / BIOS	Windows 10 Base	OS1
File Name	bm_E1VR1B2OS1	

Test Period

Start time	10:09:00	Duration	0:04:22
Stop time	10:13:22	Room Temp	

Energy Use

	Energy (Wh)	Power Factor	Voltage
System			

	Minimum	Maximum	Average
Power (W)			

Performance

	Average (fps)	Min (fps)	Max (fps)	99th% time (ms)	Dropped (%)	Runt (%)
FCAT						
FRAPS						
PresentMon						

Other

	Average	Min	Max
Sound			
CPU Temp			
GPU Temp			

All game systems testing will be conducted by the LBNL Workstation Support Group (WSG), which has a workforce of qualified technical staff with work experience on computer software and hardware systems. In addition, the WSG facilities are properly equipped to accommodate the computer systems component level hardware assembly and disassembly work that this testing will require. In order to facilitate the large logistical burden of conducting the over 1000 anticipated tests, the project has implemented a Google Documents based system of project management forms to organize the large volume of Testing Work Orders and Testing Results Reporting that will occur. The system automates both the gaming system configuration specifications for any given WSG Tester and the logging of each test session result into a common project database.

CA Gaming Systems - Nameplate Test Input Sheet

Description	
Notes	

Technician	Leo Rainer
LETS Hours	1.2
Date	11/30/2016

	Description	ID
System	HP Pavillion All-in-one Base	2
Component	GPU	
Model	GeForce GTX 1070	
Nameplate Power	250	

Data

Test	Idle (average)	Idle (max)	Stress (average)	Stress (max)
V1 (volts)				
A1 (amps)				
P1 (watts)				
V2 (volts)				
A2 (amps)				
P2 (watts)				
V3 (volts)				
A3 (amps)				
P3 (watts)				